

***INNOVATION IN THE ANALYSIS OF  
THERAPEUTIC CHANGE:  
COMBINING BOTH IDIOGRAPHIC AND  
NOMOTHETIC APPROACHES IN ONE VISUAL  
ANALYSIS.***

---

Neville M Blampied  
University of Canterbury  
Christchurch, NZ



# The Challenge of Change

Change central to all applied & clinical psychology

Detect

Induce

Measure

Analyse

Theorise

Explain

Predict & Control

**CHANGE**

# Methodological challenges researching change (1)

## Focus

- within-participant **change**
- Not between participant **difference**
  - Rutherford “stamp collecting”
  - Lakens (2013)  
“designism”

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, 4, article 863

*It depends on whether we are interested in causation at the group or individual level. If we are interested in causation at the group level, obtaining differences in between group means is reasonable. But if the causal hypothesis is at the individual level, obtaining differences between group means might be fine for a start, but it is not the basis for a strong conclusion. ...*

Trafimow, D. (2016). The mean is a multilevel issue. *Frontiers in Psychology*, 5, Article 180

# Methodological challenges researching change (2)

## Focus

- On Individuals, not
- Group means

*The application of knowledge is always to the individual case (Allport, 1942, p 58)*

*Causality operates on single instances, not on populations. (Cohen, 1994, p 1001).*

So we need

**idiographic**

as well as

**nomothetic**

science

*... it is the individual organism that is the principle unit of analysis in the science of psychology. (Barlow & Nock, 2009, p19).*

# Nomothetic vs idiographic science

## *nomothetic*

- Concerned with general laws
- Concerned with the universal
- Abstract
- Timeless
- Objective/impersonal
- Inter-individual research
- Legacy of Quetelet/Fisher

## *idiographic*

- Concerned with the individual case in context
- Concerned with the particular
- Concrete
- Historically situated
- Subjective/personal
- Intra-individual research
- Legacy of Bernard/Pavlov/Skinner

# This is not a trivial issue

Addictive Behaviors 36 (2011) 1228–1232



Contents lists available at ScienceDirect

Addictive Behaviors



## Utilizing reliable and clinically significant change criteria to assess for the development of depression during smoking cessation treatment: The importance of tracking idiographic change ☆

Andrew M. Busch <sup>a,b,\*</sup>, Theodore L. Wagener <sup>a,b</sup>, Kristin L. Gregor <sup>a,c,1</sup>, Kevin T. Ring <sup>c</sup>, Belinda Borrelli <sup>a,b</sup>

<sup>a</sup> Alpert Medical School of Brown University, United States

<sup>b</sup> Centers for Behavioral and Preventive Medicine, Program in Nicotine and Tobacco, The Miriam Hospital, United States

<sup>c</sup> Centers for Behavioral and Preventive Medicine, The Miriam Hospital, United States

### ARTICLE INFO

#### Keywords:

Smoking  
Depression  
Reliable change  
Clinically significant change

### ABSTRACT

Studies typically measure mood changes during smoking cessation treatment in two ways: (a) by tracking mean change in depression scores or (b) by tracking the incidence of major depression development using diagnostic assessments. However, tracking mean change does not capture variability in individual mood trajectories, and diagnosing participants at multiple time points is time and labor intensive. The current study proposes a method of assessing meaningful increases in depression without the use of diagnostic assessments by utilizing reliable and clinically significant change criteria. This method was applied to 212 participants in a smoking cessation trial to explore the relationship between smoking status and depressed mood, assessed at baseline, end-of-treatment, and 2-, 6-, and 12-month follow-ups. High rates of reliable (24–28%) and both reliable and clinically significant increases (23–24%) in depressed mood were observed across all participants, regardless of whether or not they achieved abstinence. However, when we calculated group mean change in depression during the trial, only decreases in depressed mood were observed across several intervals. Findings indicate that utilizing reliable and clinically significant change criteria to track symptoms of depression during smoking cessation treatment leads to different conclusions than simply tracking mean changes. We propose that a combination of reliable and clinically significant change criteria may serve as a useful proxy measure for the development of major depressive disorder during smoking cessation.

# Investigating change –

## Some questions -

Is there a better/more useful way to do it?

Is there an answer to Barlow's question?

*Why cant we be more idiographic in our research?*

Barlow & Nock (2009). *Perspectives on Psychological Science*, 4, 19 – 21.

Can we combine nomothetic & idiographic methods of analysis?

Can we incorporate visual analysis?

*As soon as you have collected your data, before you compute any statistics, **look at your data.***

Wilkinson & Task Force (1999) *American Psychologist*, 54, 594 – 604 (emphasis in original)

Can we integrate with *the new statistics*?



# *The new statistics*

- Estimation }
  - Precision – }\*
    - Confidence intervals
  - Effect sizes (descriptive)
  - Meta-analysis (inference)  
(best evidence synthesis)
- \*Both lead to concern for  
Measurement: validity/reliability/error

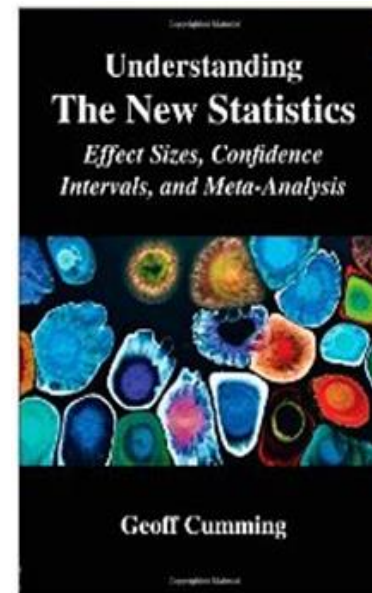


- Does not routinely use NHST

... *friends do not let friends compute p* [Klein, 2013].

*I conclude from the arguments and evidence I have reviewed that*

*best research practice is not to use NHST at all* [Cumming, 2012]



# Beginning at the beginning –

# Brinley Plots - 1965

Experimental Cognitive  
research

Brinley's insight

Scatterplots –

Plot same DV on X & Y

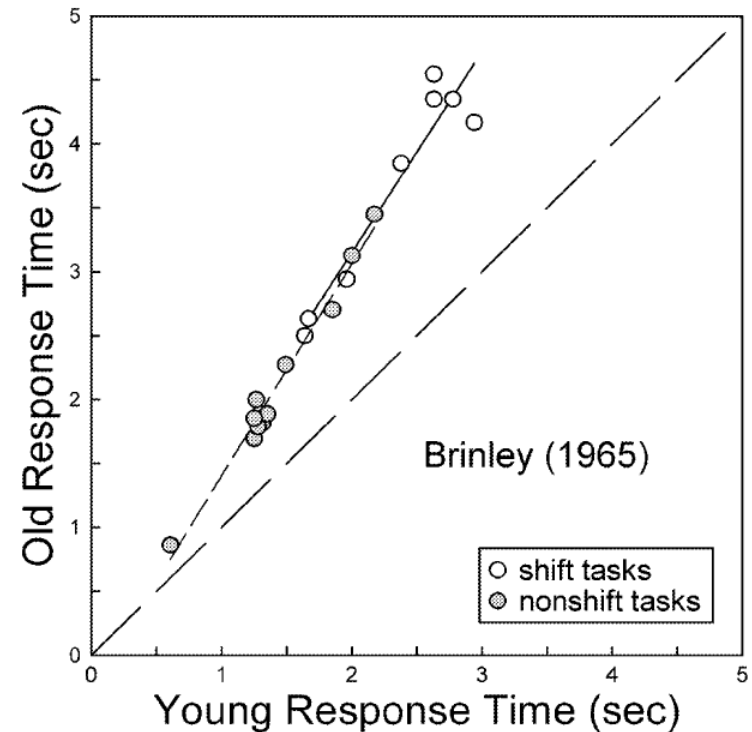
If

- X scale = Y scale, &
- Common origin
- Then  $45^\circ$  diagonal =  
line of no effect/ $X = Y$

Systematic effects deviate from  
the line

# Brinley Plots – the original

Coordinate pairs plotted represent sub-group means grouped by categorical variable (age)

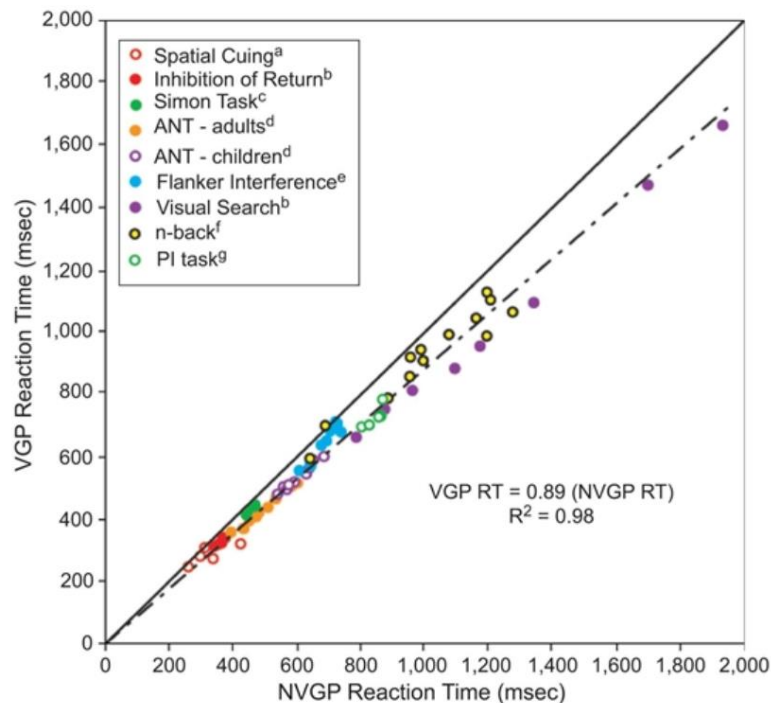


Brinley, J.F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In A.T. Welford & J.E. Birren (Eds.). *Behavior, ageing, and the nervous system* (pp 114 – 149). Springfield, IL: Charles C. Thomas.

Fig 1 from Myerson, J, Adams, D.R., Hale, S., and Jenkins, L. (2003). Analysis of group differences in processing speed: Brinley plots, Q-Q plots, and other conspiracies. *Psychonomic Bulletin & Review*, 10, 224 – 237.

Figure 1. Mean response time (RT) of the older adult group as a function of the mean RT of the young adult group in the corresponding experimental condition. The solid line is fit to the data from the 9 shift conditions (open circles) and the dashed line is fit to the data from the 12 nonshift conditions. If the condition mean RTs for the old and young groups were equal, the points would fall along the diagonal. Data are taken from Brinley (1965).

**Contemporary Example:** Dye, Green, & Bavelier, (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science*, 18 ( 6), p321-326



**Fig. 1.** A Brinley plot showing the reaction time (RT) of non-video-game players (NVGPs) on the X-axis versus that of expert video-game players (VGPs) on the Y-axis, for 39 different experimental conditions from nine different types of task. For each experimental condition, the RTs of VGPs and NVGPs were retrieved and plotted as one separate data point. A simple linear function ( $y = mx$ ) was used to describe the relationship between VGP and NVGP RTs (dashed line). VGPs responded 11% faster than NVGPs across a wide range of RTs ( $VGP\ RTs = .89 \times NVGP\ RTs$ ,  $R^2 = 0.98$ ). Importantly, similar accuracy was observed across groups, ruling out an explanation in terms of simple speed-accuracy trade-off ( $VGP\ accuracy = 0.99 \times NVGP\ accuracy$ ,  $R^2 = 0.92$ ). The studies are (a) Greenfield, deWinstanley, Kilpatrick, & Kaye (1994); (b) Castel, Pratt, & Drummond (2005); (c) Bialystok (2006); (d) Dye, Green, & Bavelier (2009); (e) Green & Bavelier (2003); (f & g) Bavelier & Bailey (2007).

# Modified Brinley plots - 1979

Clinical research

Brinley plot modified

- **Individual's** data points  
@  $t_1$  (X-axis) plotted  
against  $t_2$  (Y-axis)

-

# First use in clinical context - 1979

- $t_1$  = weight change @ end of treatment
- $t_2$  = weight change @ 5 yr follow-up
- Stable weight = points on the line (NB: reverse direction on axis scale)

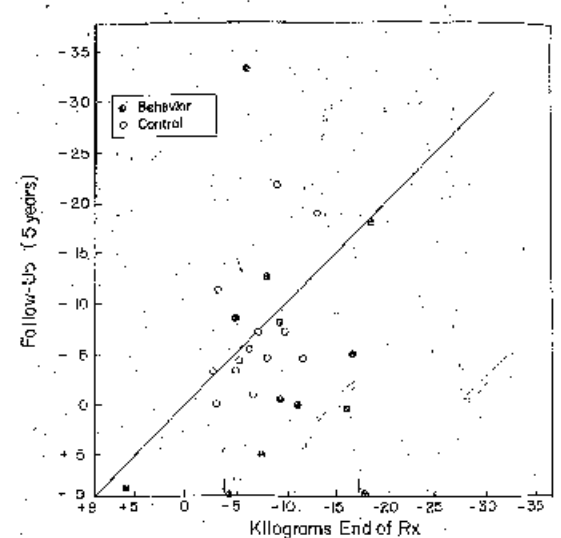


Fig 2.—Weight changes from end of treatment to five-year follow-up.

Stunkard, A.J., & Penick, S.B. (1979). Behavior modification in the treatment of obesity. *Archives of General Psychiatry*, 36, 801 – 806.

# Other clinical examples 1984 - 1995

## Therapy for alcohol abuse

- Sobell, M.B., Sobell, L.C., & Gavin, D.R. (1995). Portraying alcohol treatment outcomes: Different yardsticks of success. *Behavior Therapy*, 26, 643 – 669.

## CBT for marital distress

- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and **evaluating clinical significance**. *Behavior Therapy*, 15, 336 – 352.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to **defining meaningful change in psychotherapy research**. *Journal of Consulting & Clinical Psychology*, 59, 12 – 19.



# Rationale

Consistently  
emphasised that group  
mean data does not  
necessarily apply to any  
specific individual

Clinicians need to know  
about individual  
patterns of response to  
treatment

*...how many clients actually show  
improvement from pre- to post-  
treatment? ... Statistical group  
outcome reports convey very little  
about what types of individual  
change are typical*

Sobell, et al. (1995, p658 – 659)

# The classical approach

## CBT for marital distress

Now regarded as the *classical* approach



- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and **evaluating clinical significance**. *Behavior Therapy*, 15, 336 – 352.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to **defining meaningful change in psychotherapy research**. *Journal of Consulting & Clinical Psychology*, 59, 12 – 19.

# Parallel development in biometrics

Issue –

How to compare gold-standard assay with new assay

Lin (1989)

- Same plot
- Same logic
- $p_c = \text{concordance correlation coefficient}$

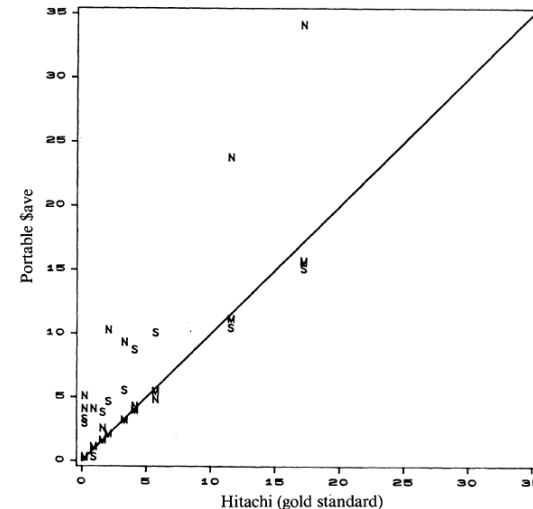


Figure 4. Portable Sava versus Hitachi by three operators for measuring bilirubin in blood.  
M: M.T.; N: Nurse; S: S-VP

# The classical approach

- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984).  
Psychotherapy outcome research:  
Methods for reporting variability and  
**evaluating clinical significance**.  
*Behavior Therapy*, 15, 336 – 352.
- Jacobson, N.S., & Truax, P. (1991).  
Clinical significance: A statistical  
approach to **defining meaningful  
change in psychotherapy research**.  
*Journal of Consulting & Clinical  
Psychology*, 59, 12 – 19.

I'm following & extending  
this work

# There is a moral to the story

This innovative research  
had almost no impact on  
subsequent clinical  
research!

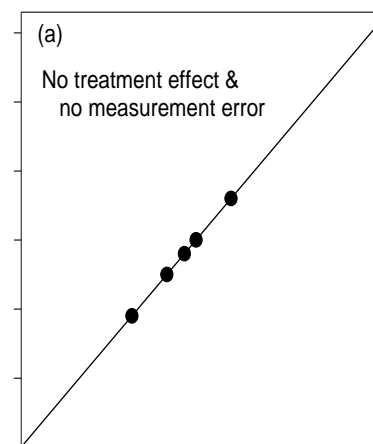
No RCT published in these  
journals has ever used  
these plots.

- Stunkard, A.J., & Penick, S.B. (1979). Behavior modification in the treatment of obesity. *Archives of General Psychiatry*, 36, 801 – 806.
- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and **evaluating clinical significance**. *Behavior Therapy*, 15, 336 – 352.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to **defining meaningful change in psychotherapy research**. *Journal of Consulting & Clinical Psychology*, 59, 12 – 19.
- Sobell, M.B., Sobell, L.C., & Gavin, D.R. (1995). Portraying alcohol treatment outcomes: Different yardsticks of success. *Behavior Therapy*, 26, 643 – 669.

# Modified Brinley plots – key features

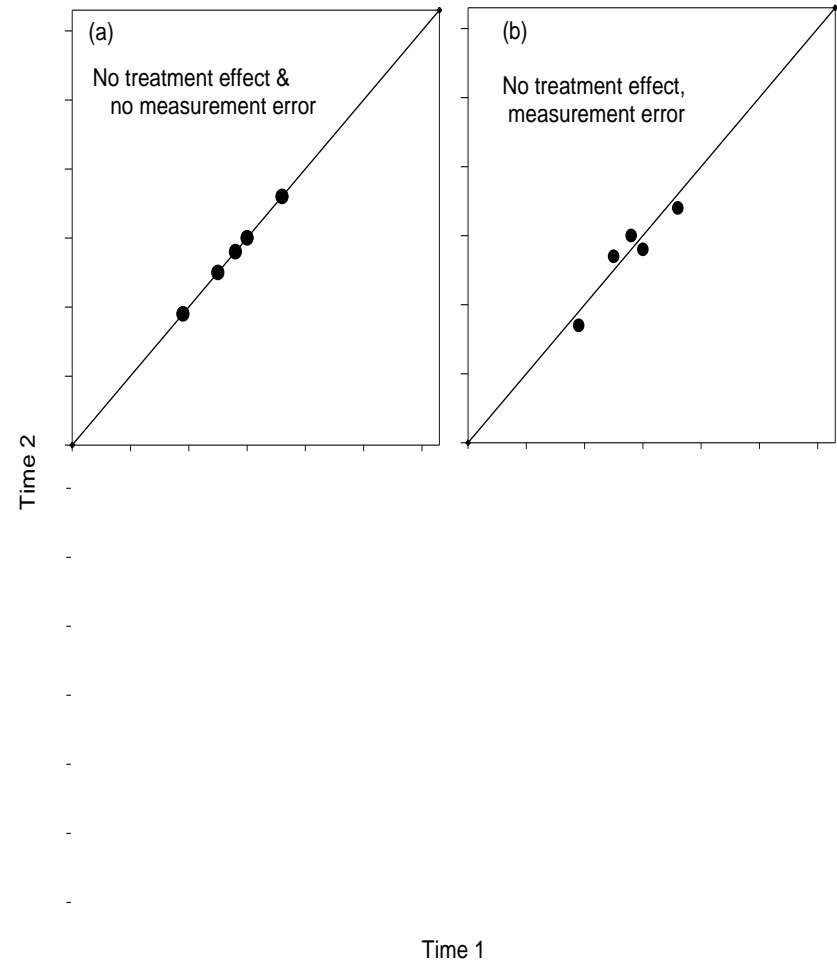
- Individual's data is plotted
- Axes same scale & origin
- 45° diagonal is line of no effect

(a) No time effect shown  
– perfect stability  $x = y$



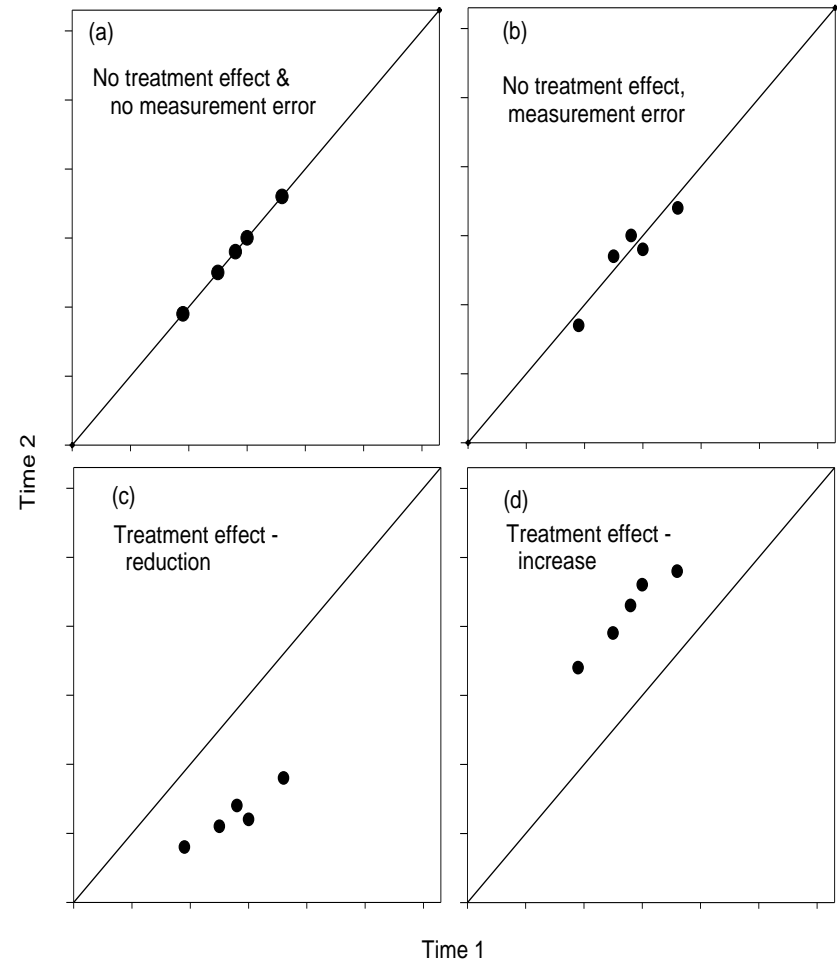
# Modified Brinley plots – key features

(b) Unsystematic variability/measurement error



# Modified Brinley plots – key features

(c & d) Systematic change over time is shown as points above/below the line





# Modified Brinley plots

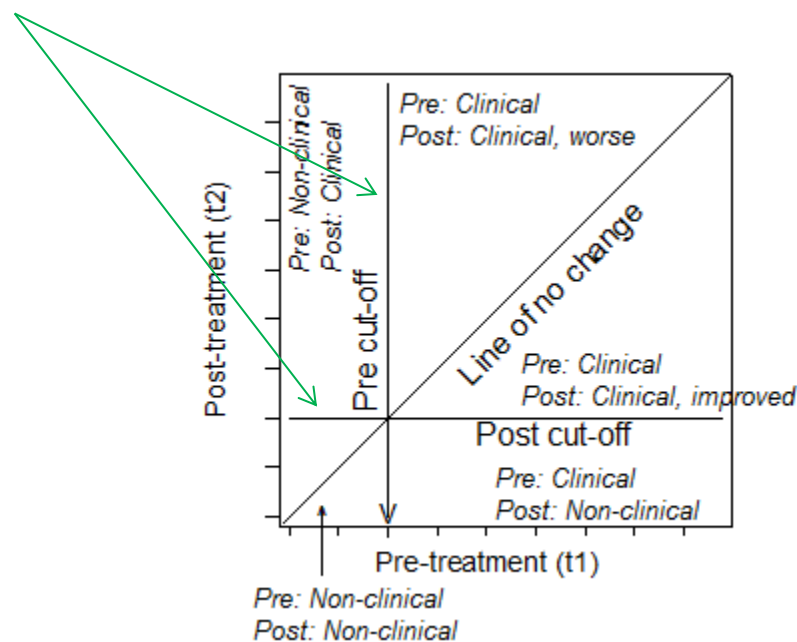
## – aids to interpretation

### Add clinical cut-off lines

- After Jacobson, et al

# Modified Brinley plots – aids to interpretation

## Clinical cut-off lines

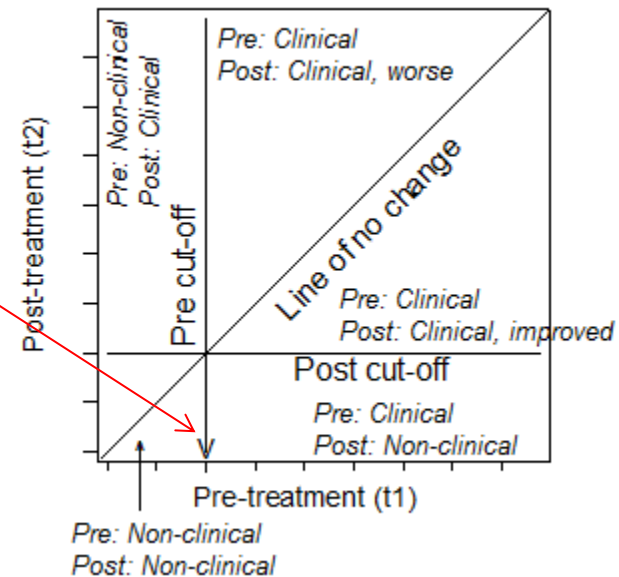


# Modified Brinley plots

## – aids to interpretation

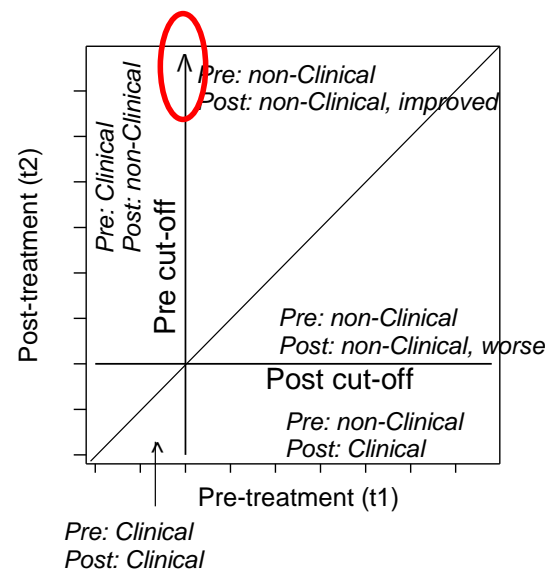
arrow indicates  
score reduction  
= improvement

➤ Graph sectors have  
meaningful  
interpretation



# Aids to interpretation ...

Interpretation where  
increase = clinical  
improvement



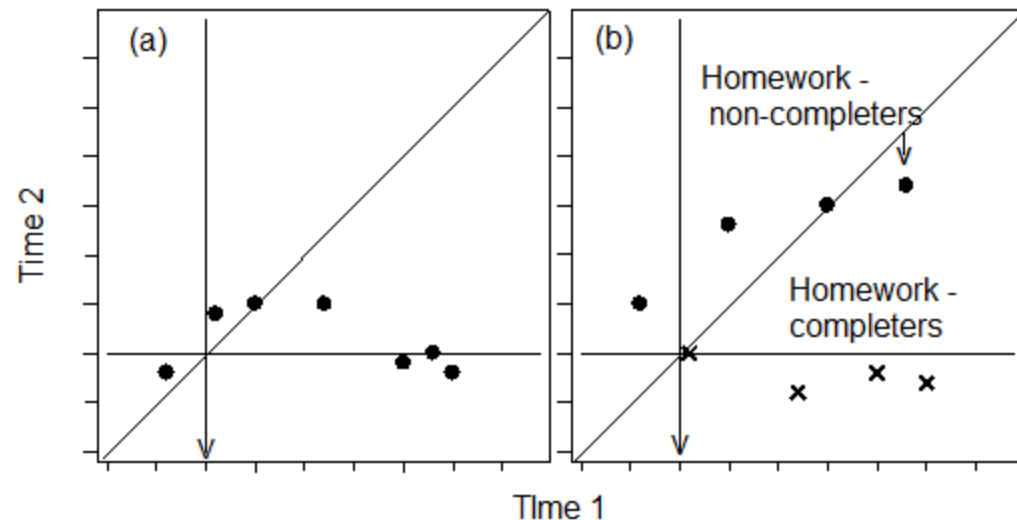
## Further Examples

- (a) Shows possible interaction between treatment and initial severity of problem

- (b) Shows effect of categorical variable split.

Many categorical variables can be investigated –

- gender
- age
- ethnicity
- therapist
- therapy features
- etc



# How much change is needed to believe that it is real change?

Suppose

- $d$  is large
- $t$  is statistically significant

Therefore group mean change is OK

But what about individual change?

How much is enough to be real?

**The Reliable Change Index**

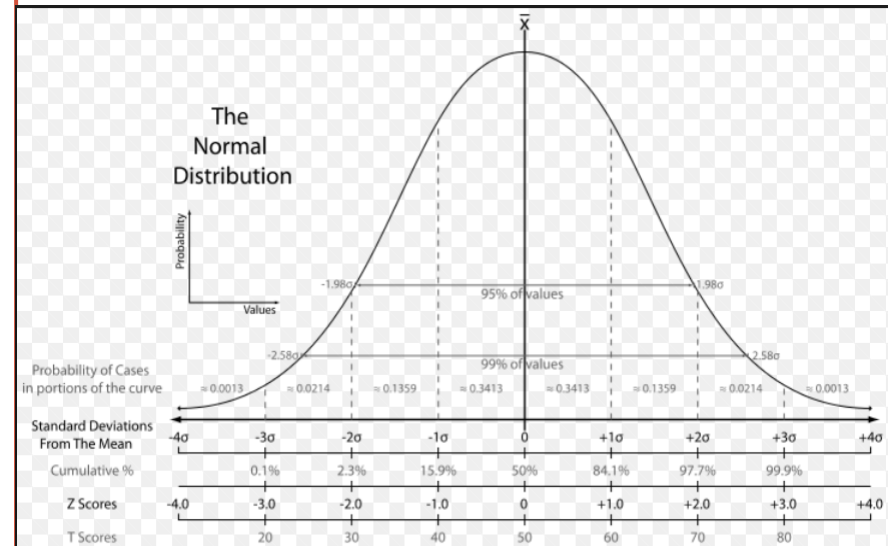
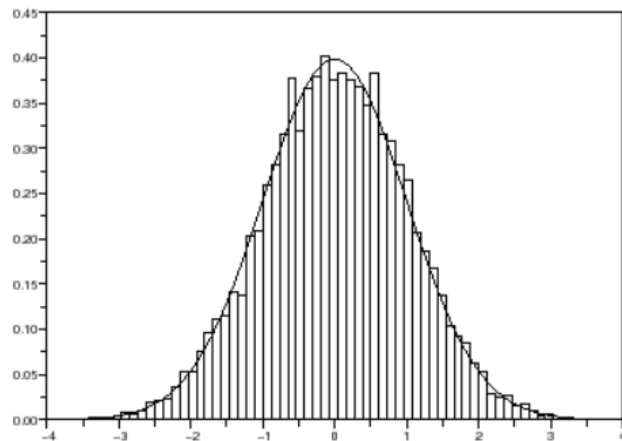
(Jacobson, et al., 1984; Jacobson & Truax, 1991)

# Reliable Change & Measurement Error

Frequency of error  
*The normal law of error*

Distribution of measurement  
error

*The Gaussian distribution*



# RC – what you need to know to compute

## Info about the measure

- $S$  = SD of reference data-set
- $r_{xx}$  = Test-retest reliability of measure (Chronbach's alpha)

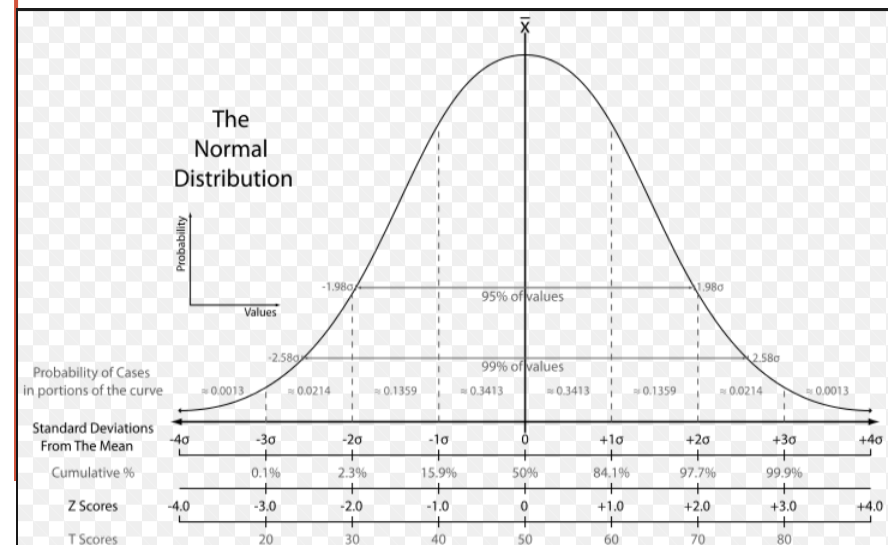
Used to compute

1. SEM
2. SDIFF

Both are a form of Standard Deviation  
SDIFF is SEM of the Error Distribution  
of the Difference Scores

## Distribution of measurement error

- Is a Normal distribution
- $SEM / S_{DIFF}$  is the Standard Deviation of the error distribution
- 95% of errors lie within  
+/- 1.96 SEM





# Logic of RCI

Is the same as for the  $t$ -test

Observation = true score  $\pm$  error

$H_0$  for any Difference Score = no difference (true score 1 = true score 2)  
(i.e., no treatment effect)

Therefore, if Difference Score  $\neq 0$ , must be due to error

BUT, if Standardized Difference Score  $> \pm 1.96$  this is improbable ( $p < .05$ ) under  $H_0$

Therefore we reject  $H_0$  and accept that there is a treatment effect (true score 1  $\neq$  true score 2) – in combination with error

# RC computation

## Steps & formulae

1. Compute Standard Error of Measurement

$$SE_M = s\sqrt{1-r_{xx}}$$

2. Compute  $S_{DIFF}$

$$S_{DIFF} = \sqrt{2(SE_M^2)}$$

3. Compute the difference score for each individual

$$Diff = x_1 - x_2$$

4. Compute  $x_1 - x_2 / S_{DIFF}$

5. If  $\geq 1.96$  a difference that large not likely due to measurement error – is in 5% tail of error distribution

**Change not likely due to measurement error  $p < .05$**

## Example

1. If variability  $s = 7.5$

Chronbach's  $\alpha = .80$

$$SE_M = 7.5\sqrt{1-.8} = 3.35$$

2.  $S_{DIFF} = \sqrt{2(3.35*3.35)} =$

**4.74**

3. So if

$$x_1 = 47.75$$

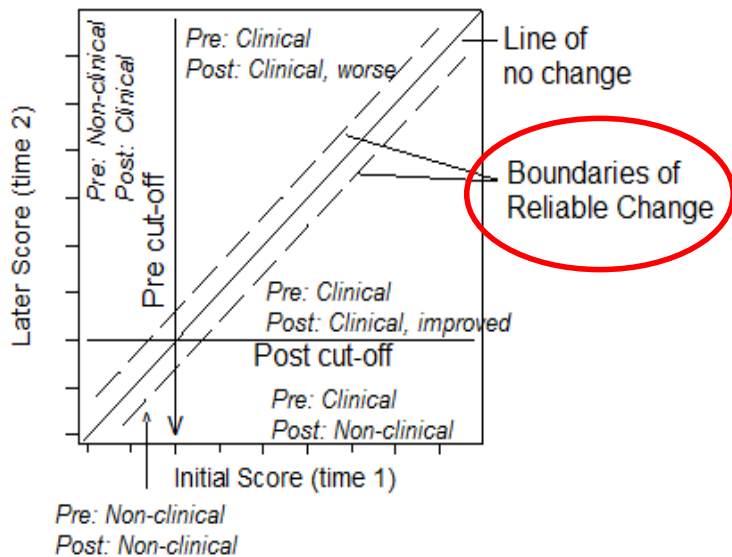
$$x_2 = 32.5 \quad Diff = 15$$

$$4. \quad 15/4.74 = 3.16$$

$3.16 > 1.96$  – **Change is reliable**

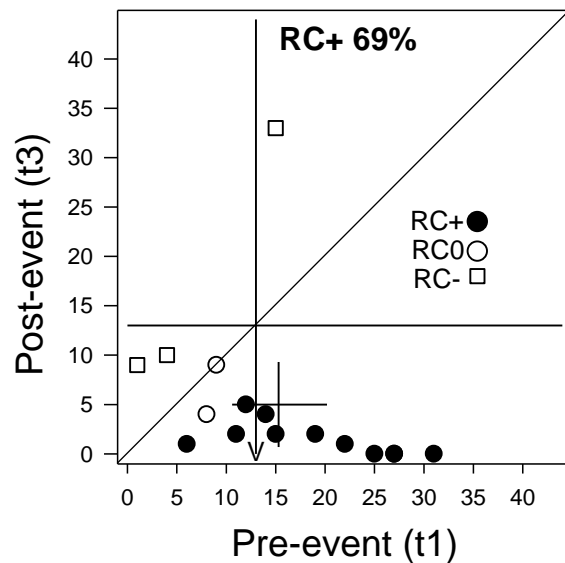
# Displaying RC information (1)

By lines



Reliable Change Index  
(boundaries) =  $\pm S_{\text{Diff}} \times 1.96$

# Displaying RC information (2)

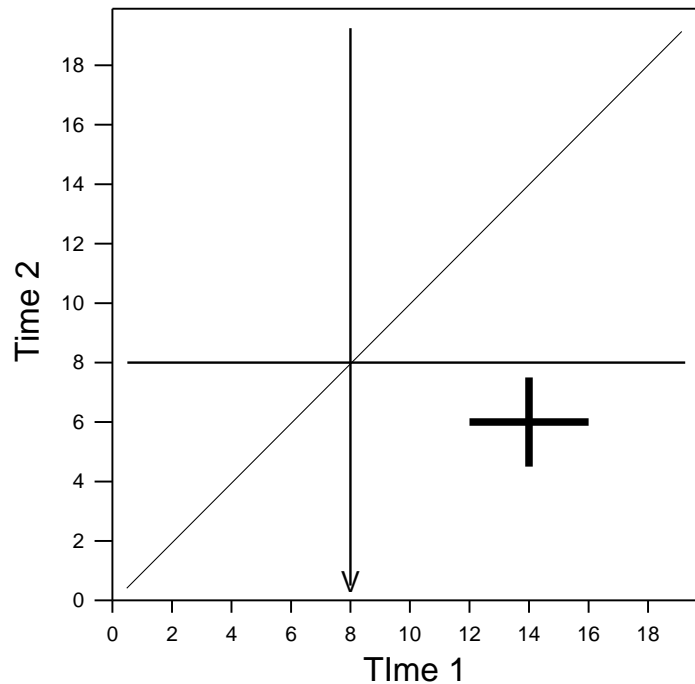


By classifying participants  
RC+, RC0, RC-

| Time 1 | Time2 | t1 - t2 | RC |
|--------|-------|---------|----|
| 31     | 0     | 31      | +  |
| 27     | 0     | 27      | +  |
| 27     | 0     | 27      | +  |
| 25     | 0     | 25      | +  |
| 22     | 1     | 21      | +  |
| 19     | 2     | 17      | +  |
| 15     | 2     | 13      | +  |
| 14     | 4     | 10      | +  |
| 11     | 2     | 9       | +  |
| 12     | 5     | 7       | +  |
| 6      | 1     | 5       | +  |
| 9      | 9     | 0       | o  |
| 7      | 4     | 3       | o  |
| 4      | 10    | -6      | -  |
| 1      | 9     | -8      | -  |
| 15     | 33    | -18     | -  |

# Interpreting the plot ... Overlaying Nomothetic Group information

Means @  $t_1$  &  $t_2$

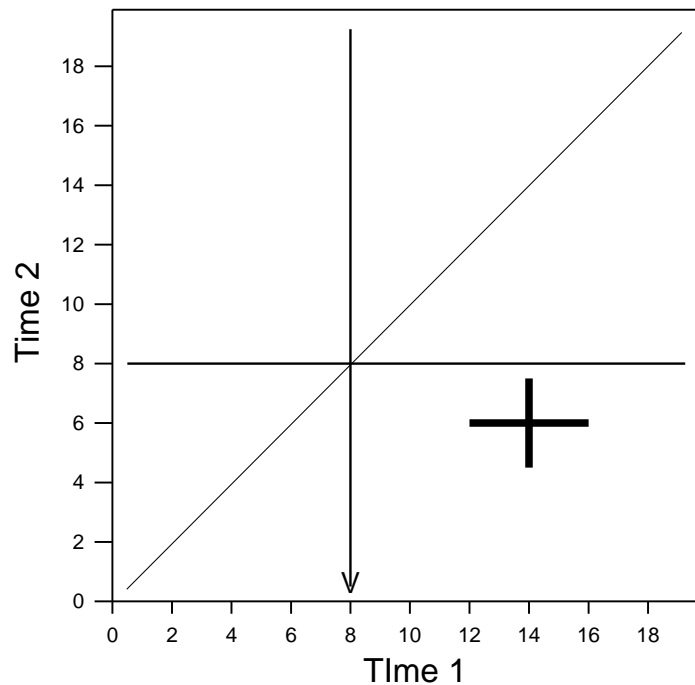


+ marks the means

# Interpreting the plot ...

## Confidence Intervals

Means  $\pm$  95% CI



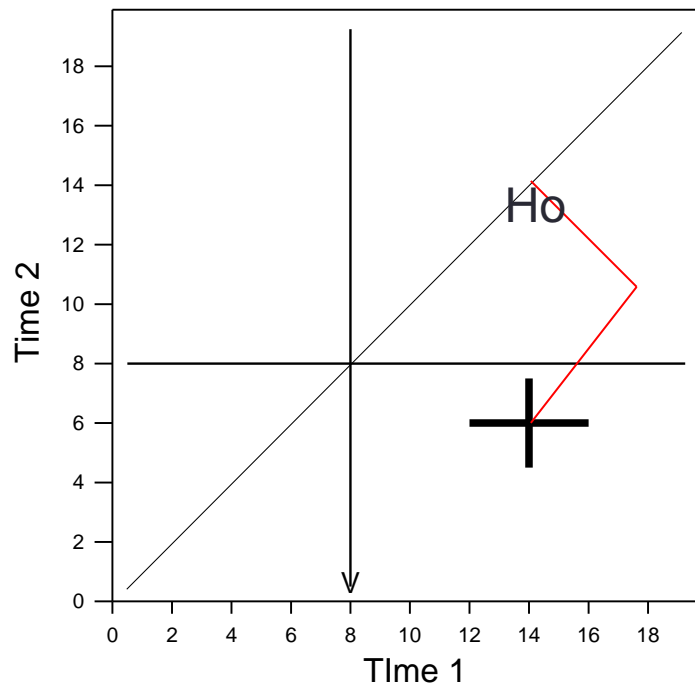
What does the CI mean?

*The interval [95% CI] estimates  $\mu$ , with 95% confidence.*

[Klein, R.B. (2013) *Beyond significance testing* 2<sup>nd</sup> Ed. p 41]

# Interpreting the plot ... relation to $t$ -test

$t$ -test



Null hypothesis is that the point lies on the diagonal

Mean 1 = Mean 2

Gap is tested by:

- Repeated measures  $t$
- 95%CI on difference scores

# Interpreting the plot ...

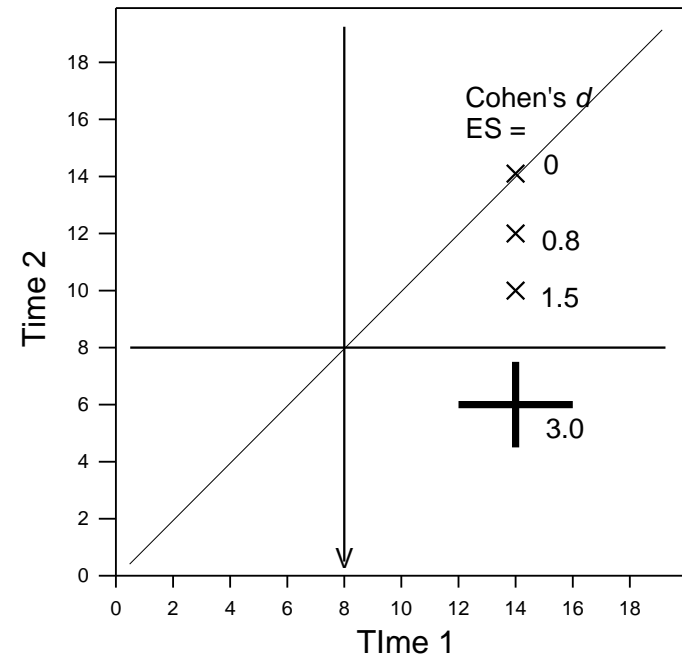
## Cohen's $d_{av}$

For within-subjects use  $d_{av}$   
(or  $d_{rm}$  to control for correlation)

For calculation of  $d_{av}$  etc  
See Cumming (2012) &  
Lakens (2013)

[both provide free software apps  
Cumming's app also calculates the  
95% CI on  $d$ ]

### Cohen's $d_{av}$ Effect size





# RC+% - a new Effect Size

- $RC+\% = nRC+/N \times 100$
- Is another **Effect Size** measure
  - Measures individual impact
    - Vs
      - Cohen's *d*<sub>av</sub>
  - Measures group mean impact

# Another useful ES

Common Language Effect Size CLES =

*The probability of any case having a better score at time 2 than time 1 (Lakens, 2013)*

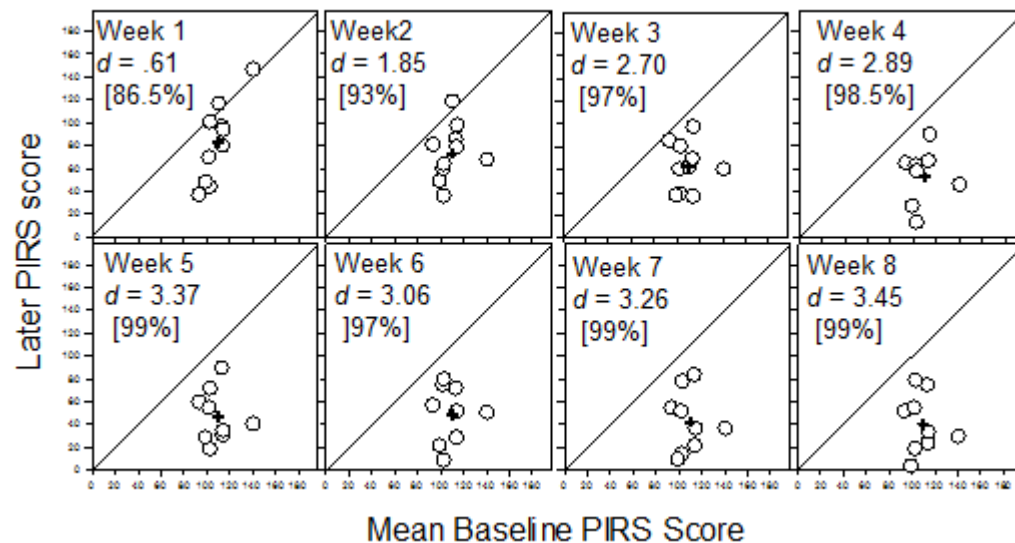
Aka Percent superiority ES

[McGaw, K.O., & Wong, S.P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111, 361 – 365]

NB: RC+% is more conservative – requires reliable change + shift in clinical direction

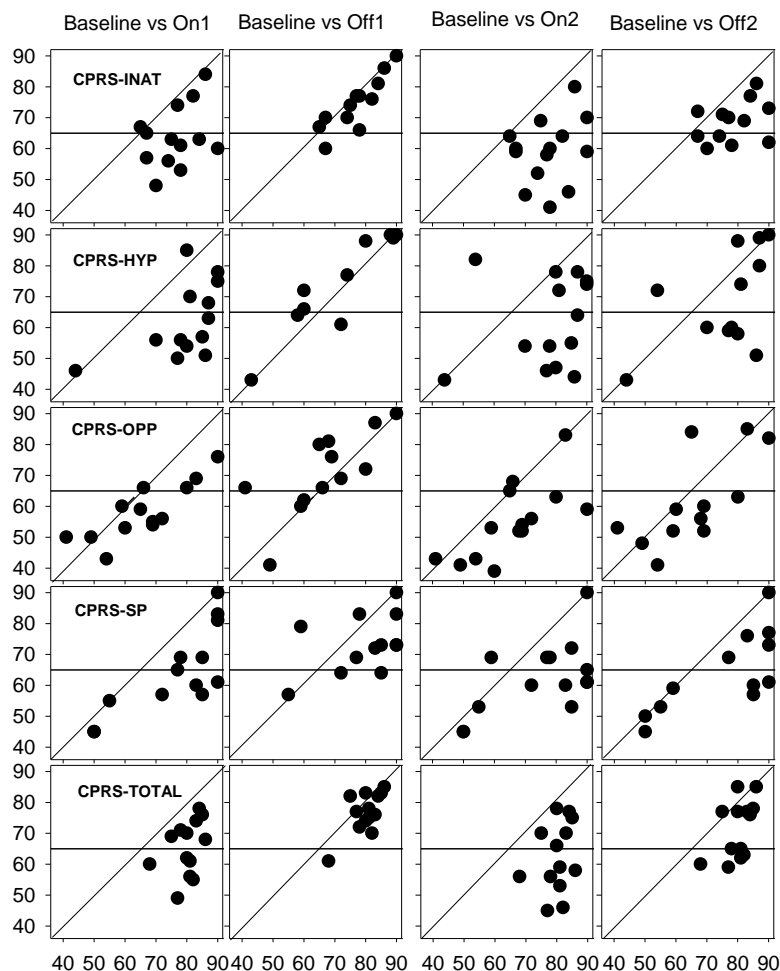
# Uses: Tracking change over time

- Group mean +
- $ES1 = d_{rm}$
- $ES2 = CLES$



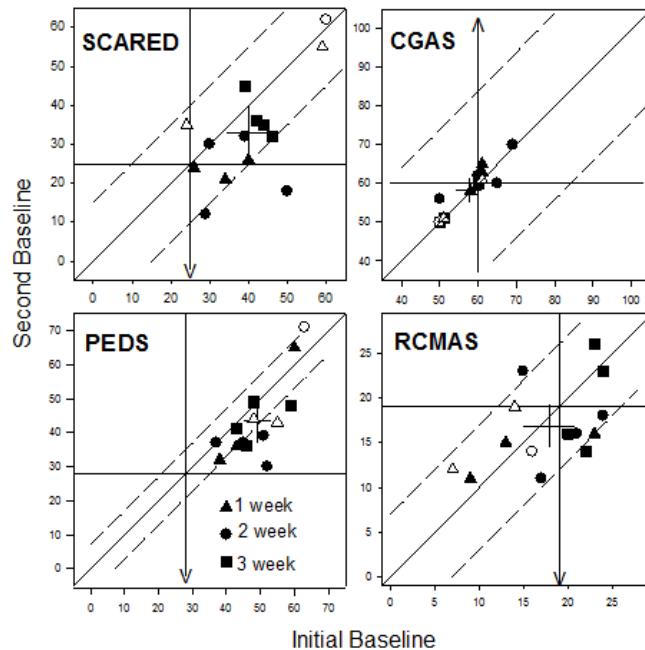
(data from Lothian,  
Blampied, & Rucklidge,  
*Clinical Psychological  
Science*, 2016)

# Uses: Summarising much data



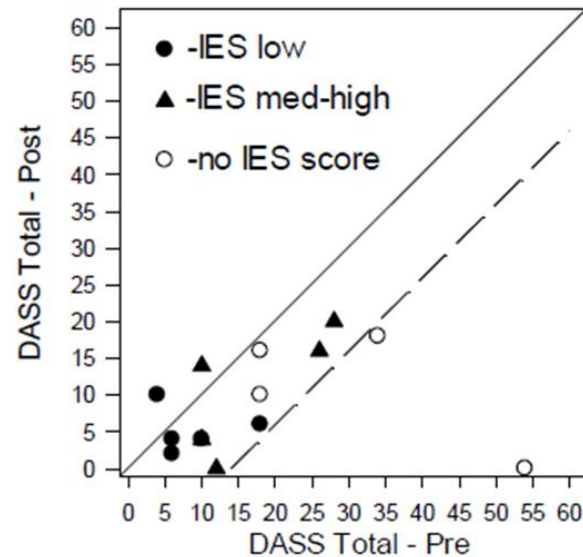
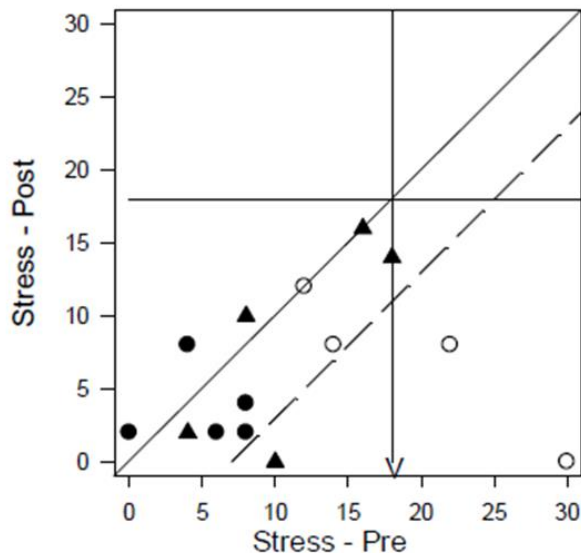
14 subjects x 5 phases x 5 DVs  
 = 350 data points  
 (Gordon, et al., 2015)

# Uses: Confirming that baselines are stable



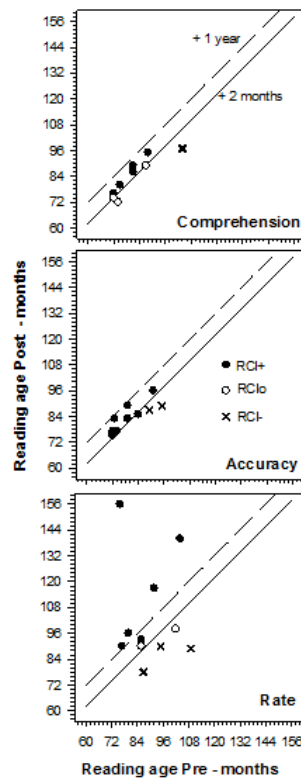
[Sole, et al., 2016]

# Uses: Differentiating participants



From  
Norton,  
Blampied, &  
France,  
2016 -  
Use of Teen  
Triple-P post  
Earthquake  
IES = Impact  
of Event  
Scale

# Uses: Adapting for absolute changes over time and practice effects



- Solid diagonal shifted upwards represents

$$Y = X(\text{age}) + 2\text{mo}$$

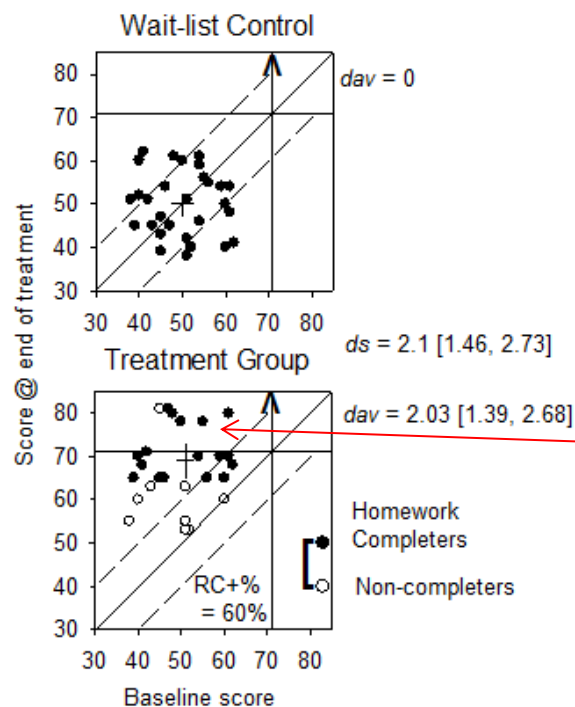
(Reading age)

- Dashed diagonal represents

$$Y = X(\text{age}) + 12\text{mo}$$

Robson, Blampied & Walker (2015)

# Uses: Revealing what conventional analyses conceal

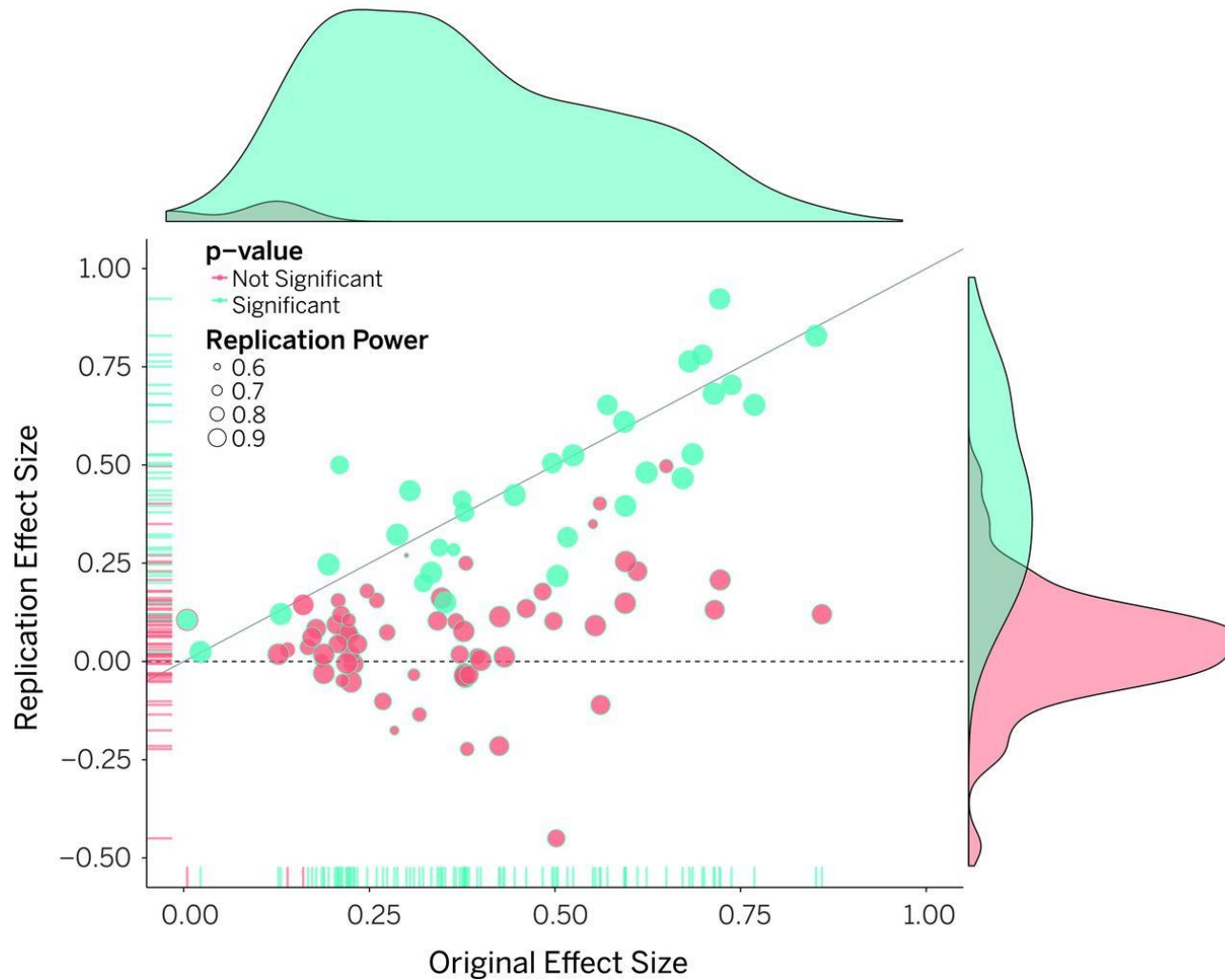


- Between & within-groups NHST statistically significant
- ES large
- **BUT – few participants show clinically significant change**

[data courtesy of Rosemary Tannock]



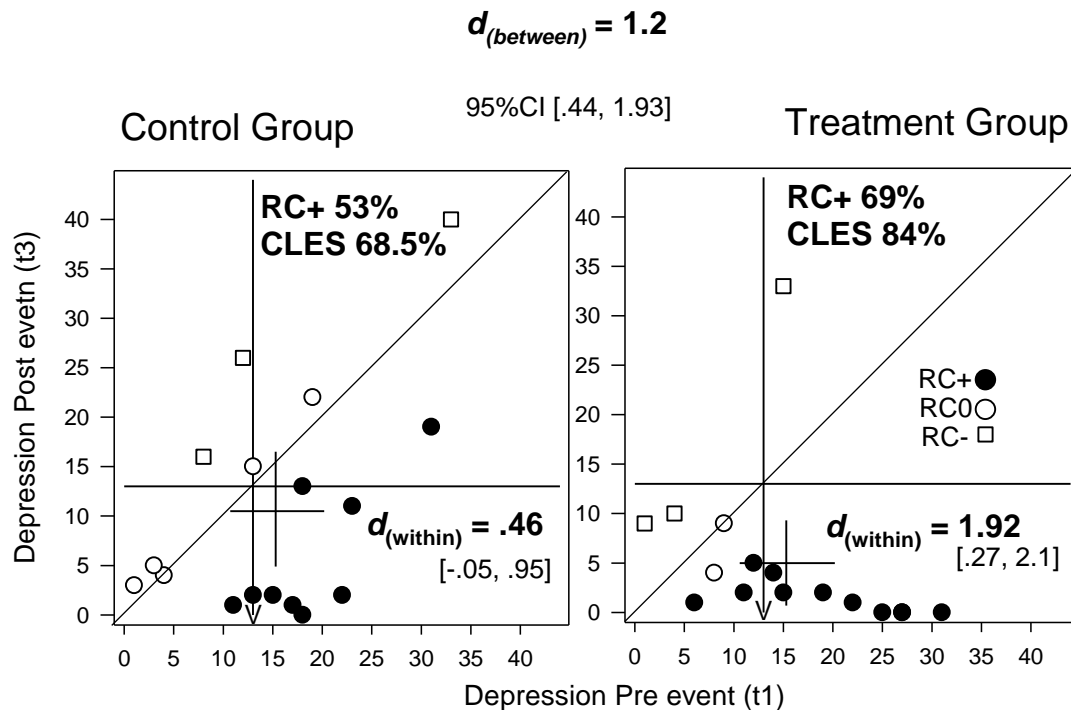
## USES: Original study effect size versus replication effect size (correlation coefficients).



Open Science Collaboration Science 2015;349:aac4716  
<http://www.sciencemag.org/content/349/6251/aac4716>



# One graph to rule them all?



- Visual analysis ✓
- Groups ✓
- Means ✓
  - 95% Confidence intervals ✓
- Individuals ✓
- Reliable Change ✓
- RCI(p < .05)
- Effect size ✓
  - 95% CI on ES ✓
- % with Reliable Change ✓
- Clinical significance ✓

## Conclusion: Potential synthesis merging nomothetic & idiographic research & the new statistics?

The new statistics<sub>(nomothetic)</sub> + single-case/  
individuals<sub>(idiographic)</sub> + replication<sub>(both)</sub> + visual analysis<sub>(both)</sub>

Abandon our over-reliance on NHST  $p < .05$

And

- Show proper respect for measurement – Calibration & SEM
- Use Reliable Change @ individual level
- Attend to clinical/practical significance [Effect size] instead of statistical significance

*Humans, not the gods, created all forms of enquiry, and we can and should modify them.*  
[Camic, Rhodes, & Yardley, 2003, p4]

# Acknowledgements & Thanks

## Colleagues

Julia Rucklidge  
Karyn France  
Jacki Henderson  
Lawrence Walker  
Mark Sobell  
Bill Fordyce

## Students – for letting me play with their data

Jo Capstick  
Amy Clarke  
Heather Gordon  
Martin Hood  
Wendy Kelly (VUW)  
James Maguire  
Jess McIvor (Massey)  
Joanna Lothian  
Pip Newton  
Tabitha Norton  
Cathy Robson  
Ellen Sole  
Frances Sutherland  
Richard Straight  
Vic Weedon (Massey)

# Abstract

For decades there have been calls for clinical research in psychology to be more idiographic and less dependent on group statistical inference, because what applies in aggregate (nomothetic research) does not necessarily apply to any specific individual (idiographic application). Recommended alternatives include more extensive use of graphs and visual analysis of data. This presentation describes the history, construction and interpretation of modified Brinley plots, a technique for analysing treatment outcomes for individuals within groups that is particularly suitable for therapy outcome research, especially during the treatment-development phase when full randomized controlled trials may be premature. Modified Brinley plots are scatter-plots that compare individual scores at time 1 (normally pre-treatment) with scores at various times post-treatment. If the origin and axis scales of the graph are the same no or little change is shown by data points clustering on or about the 45° diagonal line. Change over time (improvement or deterioration) is shown by shifts away from the diagonal. Interpretation is aided by the addition of clinical cut-offs, and by the use of the Reliable Change Index (based on measurement error), features which partition the graph space into meaningful zones. In addition to displaying individuals' data, these graphs may also display group effects such as means, variances, confidence intervals, and effect sizes. Both between-group and within-group data may be presented and analysed this way and large amounts of data can be efficiently presented and clearly understood within one figure.